ARMY RESEARCH LABORATORY

Effects of Cue Reliability on Target Detection and Visual Scanning

by Timothy L. White and James A. Davis

ARL-TR-5410 **December 2010**

NOTICES

Disclaimers

Army Research Laboratory

Aberdeen Proving Ground, MD 21005-5425

ARL-TR-5410 **December 2010**

Effects of Cue Reliability on Target Detection and Visual Scanning

Timothy L. White and James A. Davis
Human Research and Engineering Directorate, ARL

REPORT DOCUMENTATION PAGE	Form Approved OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
December 2010

2. REPORT TYPE
Final

3. DATES COVERED (From - To)
August 2008

4. TITLE AND SUBTITLE
Effects of Cue Reliability on Target Detection and Visual Scanning

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Timothy L. White and James A. Davis

5d. PROJECT NUMBER
62716AH70

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
U.S. Army Research Laboratory
ATTN: RDRL-HRS-B
Aberdeen Proving Ground, MD 21005-5425

8. PERFORMING ORGANIZATION REPORT NUMBER
ARL-TR-5410

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Past research has indicated that automated target acquisition systems augment target detection in a way that facilitates faster detection times and increased target detection. However, in real-world applications, such target acquisition systems provide imperfect target classification and are prone to false positive and false negative errors when attempting to identify targets in operational environments. Furthermore, past research indicates a direct relationship between the reliability of the target acquisition system and performance on the target detection task. Few studies address the effect of cue format and target salience on task performance with respect to reliability level. This experiment examined the effect of four reliability levels on task performance (0% reliability or baseline, 60% reliability, 75% reliability, and 90% reliability) and highlighted the importance of cueing context on cueing effectiveness. For this experiment, the salience of the cues paired with a nonconducive cue format produced an inverse relationship between reliability and task performance. Overdependence on these cues resulted in target detection times that actually increased with increasing reliability. Results from this experiment also showed that even infrequent audio tones (as seen in the 60% reliability condition) could potentially increase operator awareness. These results provide invaluable information regarding the contextual importance of target detection cues and importance of cue format for target acquisition systems in operational environments.

15. SUBJECT TERMS
reliability, target cueing, auditory, target detection, automated systems

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON / 19b. TELEPHONE NUMBER (*Include area code*)
Unclassified	Unclassified	Unclassified	UU	40	Timothy L. White / 410-278-5884

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

Contents

List of Figures

List of Tables

Acknowledgments

The authors would like to express their appreciation to the following individuals who assisted in the conduct of this research:

- Monica Glumm and Kathy Kehring for their assistance with the development of this research effort,
- Kelvin Oie for his assistance with the development of the virtual environment for the simulator, and
- GySgt Michael Peavey for his assistance in obtaining volunteers from the U.S. Marine Corps Detachment for this study.

Last, but never least, the authors would like to thank the U.S. Marines who participated in this research for their enthusiasm and dedication to this project.

1. Introduction

Vision provided through periscopes and other sighting devices in current armored combat vehicles often limits the ability of crews to quickly detect hazards and other threats within their local environment. Therefore, when not actively engaged in combat, vehicle commanders often prefer direct line-of-sight of their local environment (e.g., via open-hatch or name-tag position), sacrificing armor protection for an unobstructed view of the area around their vehicle. In urban terrain, however, hazards can lie at all azimuths and elevations, and open-hatch operations are particularly risky. Operational requirements for future combat systems manned ground vehicles state that Soldiers must maintain platform capability while operating in the closed-hatch position. This necessitates indirect-vision systems that provide the field-of-view and resolution to effectively perform area security and target acquisition tasks.

Analyses of the functions and tasks of crewmembers using the Improved Performance Research Integration Tool (IMPRINT) indicate that Soldiers will often experience high levels of workload when scanning for targets and performing other critical tasks concurrently (Mitchell et al., 2004; U.S. Department of Defense, 2005). In an effort to reduce the burden on overtaxed resources and enhance situational awareness and understanding, U.S. Army researchers are investigating new technologies and techniques to help Soldiers perform area security and target acquisition tasks. Although the technology needed to detect targets and provide target location cues in one's local environment is not fully mature, research examining and quantifying the advantages of such techniques is considered an important first step. The findings of Glumm et al. (2006, 2007), shown in table 1, indicate that when cues are provided about the location of targets, time to first shot can potentially be reduced by 43% and the percentage of hits increased by 55% compared to a baseline (no cue) condition for a simulated shooting task. Improvements were also found in the performance of a secondary communications task (22%), along with reductions in perceived workload (12%). For the purpose of their assessment of cueing techniques, Glumm et al. (2006, 2007) assumed that the automated system which provided the target location cues would detect 100% of the targets with no false alarms. However, it was recognized that, in the real world, it was unlikely that such a system would be totally reliable. As research has shown, decreases in the reliability of automated systems can affect their use and reduce the potential benefits they were thought to provide over nonautomated systems. Wickens and Dixon (2005) compared data points derived from 20 studies measuring performance with automated systems that varied in reliability and a baseline, nonautomated condition. Their analysis indicated that when the reliability of the automated system was about 70% or below, performance with the automated system was no better than performance with the nonautomated system.

Table 1. Percent difference in target acquisition performance and workload between baseline (no cues) and unimodal (Glumm et al., 2006) and multimodal (Glumm et al., 2007) cues about target location.

	Percent Difference From Baseline (No Cues)			
Unimodal Cues	**Time to First Shot**	**Percent Hits**	**Information Recalled**	**Overall Workload[a]**
Visual	–38	+56	+39	–19
Verbal	–38	+56	+29	–16
Three-dimensional (3-D) audio	–17	+45	+24	–6
Tactile	–29	+54	+28	–18
Mean	**–31**	**+53**	**+30**	**–15**
	Multimodal Cues			
Visual + verbal	–44	+56	+20	–11
Visual + 3-D audio	–43	+55	+26	–14
Visual + tactile	–42	+54	+24	–10
Verbal + tactile	–43	+56	+17	–13
Mean	**–43**	**+55**	**+22**	**–12**

[a] Computed from weighted ratings of mental, physical, and temporal demands, performance, effort, and frustration using the National Aeronautics and Space Administration-Task Load Index (NASA-TLX).

There are several challenges associated with incorporating automated target acquisition systems in real operational environments. One such challenge involves the operator's degree of reliance on the automated system. Studies have indicated that underutilization (disuse) and overutilization (misuse) of automation can occur regardless of the level of system reliability (Beck et al., 2000; Dzindolet et al., 2001, 2003). Automation bias (overestimation of system ability) and overly trusting the system can lead to misuse, whereas self-serving bias (overestimation of one's own ability) and overly distrusting the system can lead to disuse (Dzindolet et al., 2001, 2006). The reasons for inappropriate use of automation are many, reflecting differences among individuals as well as changes in judgment and attitudes that can occur with changes in the situation or the environment. Operators may tend to rely on an automated system as an effort-saving strategy (Mosier and Skitka, 1996) particularly during periods of high workload (Parasuraman et al., 1993) or when fatigued (Dzindolet et al., 2006).

Another challenge associated with incorporating automated target acquisition systems in real operational environments is associated with the reliability (or accuracy) of the systems. As with all predicative technologies, there is the possibility of type I (i.e., false positive) and type II (i.e., false negative) errors. The system might indicate that a target is present when it is not (i.e., a false alarm), or the system might indicate that a target is not present when it is (i.e., a miss). An automated system that identifies the location of targets may also indicate that a target is at one location when the target is at another location. All of these errors can have serious consequences. False alarms can potentially result in fratricide. Friendly lives can also be lost if targets are missed or errors made in localization. In a combat situation, if the false alarm rate of

the automated system is high, it is believed that the potential for a missed target and loss of life might dissuade Soldiers from totally ignoring the system, but delays in responding to system alerts are possible. If false alarm rates are perceived to be too high or annoying, Soldiers may attempt to disable the system. If the miss rate of an automated system is high, it is expected that Soldiers would remain highly responsive to system alerts but continue to scan for targets as if no automation were provided. Thus, an automated target detection system that is prone to misses may not yield the anticipated benefits in reduced workload and enhanced secondary task performance.

Much of the research on automated target acquisition systems is geared toward highlighting the effectiveness of the particular system and often neglects examining the reliability of an automated system (specifically, the tendency for type I errors). Just as in the previously mentioned studies (Glumm et al., 2006, 2007), much research assumes that the automated system which provided the target location cues would detect 100% of the targets with no false alarms. It is recognized that, in the real world, it is unlikely that such a system would be 100% reliable. As research has shown, decreases in the reliability of automated systems can affect their use and reduce the potential benefits they were thought to provide over nonautomated systems. In fact, Wickens and Dixon (2005) found that when the reliability of the automated system was about 70% or below, the performance with the automated system was no better than performance without the automated system. Therefore, carefully examining the effects of the reliability of an automated target detection system or technique is critical when incorporating it into operational environments. This document will describe an effort designed to quantify the effects of reliability of an automated target detection system on a target acquisition task. The effects of system reliability are often reflected in visual scanning behavior, so this effort will also monitor visual scanning (i.e., eye and head movements) (Wickens et al., 2005) behavior associated with different levels of system reliability.

2. Purpose

Quantifying the effects of the reliability of an automated target detection system will help to define design standards for such a system. The purpose of this investigation was to quantify the effects of reductions in the reliability of an automated target detection system on visual scanning behavior, target detection, and subjective workload. Three levels of system reliability were assessed: (1) 90% reliability or 10% misses, (2) 75% reliability or 25% misses, and (3) 60% reliability or 40% misses. The data obtained at these reliability levels were compared with data obtained in the baseline, nonautomated condition in which no cues were provided about the presence or location of targets.

There were three general hypotheses in this study. First, it was hypothesized that the time to detect targets would increase and the percentage of detected targets would decrease with each decrease in level of reliability, due to differences between conditions in the number of missed targets and the size of the display area to be searched in the absence of cues about target location. It was hypothesized that no significant differences would be found between the baseline condition and the 60% level of reliability on these measures.

Second, it was hypothesized that eye movements would increase with each decrease in the level of reliability due to increases in the number of targets missed by the automated system, the frequency of search, and the size of the display area to be searched in the absence of cues about target location. Also as eye movements increase, the duration and frequency of eye fixations would decrease. We expected no significant differences between the baseline condition and the 60% level of reliability on these measures.

Finally, it was hypothesized that subjective ratings of workload would increase with each decrease in the level of reliability, due to increases in the number of targets missed by the automated system, the frequency of search, and the size of the display area to be searched in the absence of cues about target location. We expected no significant differences between the baseline condition and the 60% level of reliability.

3. Methodology

3.1 Participants

Twelve male Marines from the U.S. Marine Corps Detachment at Aberdeen Proving Ground, MD, participated in this research. Participants ranged in age from 18 to 24 years (20.1 ± 0.64).* All participants had vision (corrected or uncorrected) sufficient for driving. Participants had normal hearing, as determined by an audiogram. The voluntary, fully informed consent of the persons used in this research was obtained as required by Title 32 Code of Federal Regulations (1991) and Army Regulation (AR) 70-25 (U.S. Department of the Army, 1990). The investigators adhered to the policies for the protection of human subjects as prescribed in AR 70-25.

3.2 Apparatus

3.2.1 Control Station and Target Scenario

The control station simulated the crew station within the Crew Integration and Automated Testbed Army Technology Demonstrator (CAT-ATD) depicted in figure 1 (left). The crew station consisted of a fixed-base driving simulator with three screens that provided a view in

* Mean ± standard error.

Figure 1. Crew station in CAT-ATD vehicle (left) and CAT-ATD control station simulation (right).

front of a vehicle that was traveling through a virtual environment (figure 1, right). Each screen was 17 in diagonal, with a resolution of 1280 × 1024 pixels. Each of the three screens provided a 40° horizontal view of the scene in front of the vehicle, for a total field-of-view of 120°. The three screens were positioned side by side, with the left and right screens at a 30° angle relative to the center display. The seat position of each participant was adjusted so that his eye position was referenced to a fixed, external reference point 50 cm from the plane of the center display and his back comfortably rested against the back of the seat. The image that was presented across the three screens was a simulation of the view provided by three fixed cameras mounted on a vehicle traveling at a constant speed of 5 mph through urban terrain.

The simulated environment scene included streets and intersections bordered by one- and two-story buildings. Stationary human figures (i.e., targets) were placed within the scene (see figure 2). The urban scene and human figures were generated by custom software called SimCreator (Realtime Technologies, Inc). A standard, three-speaker system (Dell, A525) was used to present the auditory cues about the location of targets in those conditions where cues are presented. The driving simulator also included a steering wheel. Participants did not use the steering wheel to drive through the scenarios because they were only required to ride through each scenario. However, a button located on the steering wheel was used by each participant to indicate that a target was detected.

Figure 2. Illustration of three simulator displays with a target at 11 o’clock.

3.2.2 Target Location Cues

Audio speech cues provided the location of a target in the simulation relative to the participant. All cues were provided verbally in a clock-type format (e.g., "2 o'clock – 2 o'clock!"). All cues were prerecorded in a female voice, were normalized, and noise reduction was applied. For each audio signal, the peak sound pressure level (SPL) through the participant's headphones was 73-dB SPL as measured through an artificial ear. A sample view of the three simulator displays is shown in figure 3. All targets appeared in the proximity of five clock positions (i.e., 10, 11, 12, 1, and 2 o'clock) within the 120° horizontal field of view in front of the simulated vehicle (figure 3).

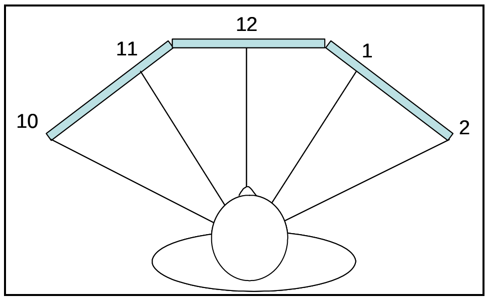

Figure 3. Clock positions within the 120° field-of-view of the three-screen display.

3.2.3 Head/Eye Tracker

The eye and head movements of each participant were tracked using a faceLAB 4* Head/Eye-Tracking System. The tracking system observed natural eye and head movement at a spatial resolution of ~0.5°. The tracker is a camera-based system that computes and logs all eye and head movement data in real time, along with data on measurement reliability. The eye and head movement data were synchronized with the target detection data from the simulation software.

3.2.4 Questionnaires

A demographics questionnaire was administered to obtain background information about each participant (appendix A). Before the experiment began, each participant was asked to complete a motion sickness questionnaire to obtain a baseline for comparison with levels of symptoms experienced during and after the experiment (Kennedy et al., 1992) (appendix B). Upon completion of each of the four conditions, a postcondition questionnaire was administered to gather each participant's opinions and comments regarding the preceding experimental condition (appendix C). The NASA-TLX with paired comparisons was used to assess the participant's experience of workload (Hart and Staveland, 1988) (appendix D). This technique uses rating scales to assess mental, physical, and temporal demands, performance, effort, and frustration.

* faceLAB 4 is a trademark of 5DT Products.

3.3 Experimental Design

Each participant completed two trials within each of the four test conditions (for a total of eight driving trials). The test conditions were baseline (no target location cues), 60% reliability, 75% reliability, and 90% reliability. The baseline condition referred to a condition when no audio cues were present. The 60%, 75%, and 90% reliability conditions referred to a condition where 60%, 75%, and 90% of the target presentations (respectively) were accompanied by an audio cue. The order of reliability conditions was counterbalanced among the participants. The duration of each trial was ~5 1/2 min. During each trial, 20 targets were presented at different locations within the urban scene and each trial was presented twice, for a total of 40 target presentations per experimental condition. All targets were stationary human figures that were revealed as the participant's vehicle moved through each trial (figure 2). From the participant's perspective, the initial distance of each target ranged from 25 to 50 m. The targets were randomly selected from a set of 10 human-figure models available within the simulation software. Each human figure had distinct characteristics that each participant identified at the time the target was detected to verify that the target had been seen. Each target was presented one at a time for 3 to 5 s. The time interval between target presentations was randomized and ranged from 5 to 20 s. The primary task of the participants was to detect and identify targets while riding through an urban terrain. Participants did not use the steering wheel to drive through the scenarios because they were only required to ride through each trial. However, a button located on the steering wheel was used by each participant to indicate that a target was detected. Participants were asked to provide a verbal description of each target detected. The investigator recorded whether or not the target descriptions were accurate. This was done to ensure that there were no erroneous button presses.

The dependent variables in this study included measures of target acquisition performance, visual scanning (eye movements), and subjective ratings of workload. In the target acquisition task, the dependent variables were target detection time and the percent of correct detections. Target detection time was computed from the time at which any portion of the target is revealed in the display to the time the participant pressed the button on the steering wheel. The percent detections measured the percentage of targets correctly detected by the participant. The dependent variables associated with visual scanning were fixation duration and fixation frequency. Mean eye fixation duration measured the average time of eye fixations during a ride, where fixations were defined as periods of at least 100 ms during which gaze position did not change more than 0.5°of visual angle (Jacob, 1993). Fixation frequency was computed by dividing the total number of fixations by the total time(s) in a given condition. Overall workload scores were computed from the weighted ratings on each of six dimensions of perceived workload (i.e., mental, physical, and temporal demands, performance, effort, and frustration) obtained with the NASA-TLX.

3.4 Procedures

3.4.1 Training Procedures

Each volunteer was briefed on the purpose of the investigation, the procedures to be followed during the study, and any risks involved in his participation. The investigator read the volunteer agreement affidavit aloud to the participant who followed along, and any questions the participant might have regarding the study were addressed. If the participant agreed to take part in the investigation, he completed the information on the last page of the affidavit and signed it. A demographic questionnaire was then administered to the participant to obtain pertinent background information. Also, a motion sickness questionnaire was completed to obtain a baseline for comparison, with levels of symptoms experienced during the study and at the conclusion of testing. The investigator then calibrated the head/eye tracker to obtain optimal eye- and head-tracking performance. After the head/eye tracker had been calibrated, the participant received instruction on how to rate his workload experience using the NASA-TLX. The participant then completed one 3-min practice ride in each of the four experimental conditions, with 2-min breaks between each ride. Prior to each ride, the participant was informed of the experimental condition he would receive. And after each ride, the participant was informed of the number of targets he did not detect during the ride. After the last training ride, the participant practiced completing the NASA-TLX to rate his workload experience.

3.4.2 Test Procedures

After a 5-min rest break, the test period began. Before each ride, the participant was informed of the experimental condition he would receive. The participant completed two rides in each of the four experimental conditions. Upon completion of each condition, the participant completed the NASA-TLX to rate his workload experience and a postcondition questionnaire to obtain comments regarding the target detection task in the condition just completed. The investigator informed the participant of the number of undetected targets in that condition. During the testing, participants were monitored for symptoms of motion sickness. At the conclusion of testing, participants completed motion sickness questionnaires to identify any elevated symptoms of motion sickness. None of the participants experienced motion sickness during this investigation.

4. Results

Linear mixed models (McCulloch, 2003) were used to analyze the effects of reliability on target detection and eye movements. Fixed factors in the linear mixed model were reliability level, condition order, and target location. Random factors were participant, participant*condition *order, and participant*condition*order*location. Post hoc evaluations were pair-wise comparisons using the least significant differences method. All statistics were reported with outliers removed. Outliers and missing data accounted for >1% of the data and were generally associated with errors. All means were reported as mean ± standard error. In addition to significant main effects, statistical trends ($0.05 < p \leq 0.10$) were also reported, as these trends might become main effects with larger sample sizes.

4.1 Detection Time

A linear mixed model revealed a significant condition effect in detection time among the four reliability conditions ($F_{3, 30} = 3.4$, $p \leq 0.04$). Pair-wise comparisons revealed that detection times during the 60% reliability condition ($p < 0.05$) were significantly faster than detection times during the baseline and 90% reliability conditions (see figure 4). Although the 75% reliability condition was also relatively faster than the baseline and 90% conditions, this difference was not statistically significant. There was no significant difference between the baseline and 90% reliability conditions.

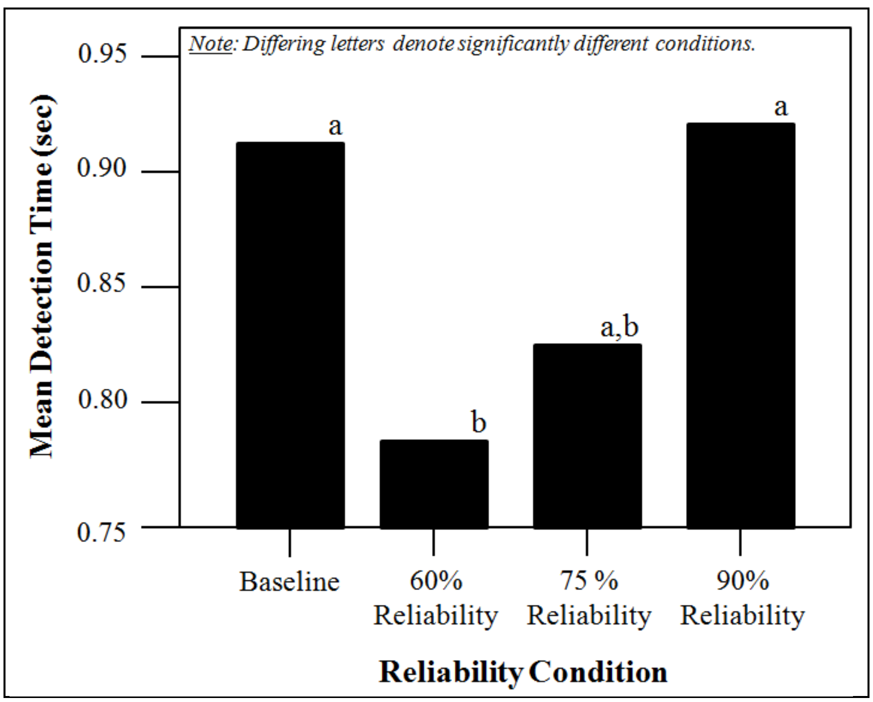

Figure 4. Detection time by reliability condition.

4.2 Percent of Detections

The linear mixed model did not reveal a significant effect of condition on the percent of detections (baseline: 97.5 ± 1.0; baseline: 97.9 ± 1.0; baseline: 99.2 ± 0.4; baseline: 98.1 ± 0.8). There was, however, an order effect ($F_{3, 30} = 7.0$, $p \leq 0.01$) where participants identified significantly less targets in their first test condition (95.6 ± 0.71) than the three subsequent conditions (99.0 ± 0.65); all p's < 0.02.

4.3 Eye Movements

A linear mixed model did not reveal any significant differences in fixation duration or fixation frequency between the different conditions.

4.4 Questionnaires

The analysis did not reveal a significant effect for reliability level on the overall weighted workload scores that were obtained from the ratings on each of the six dimensions (i.e., mental, physical and temporal demands, performance, effort, and frustration) of perceived workload. Additionally, there were no significant effects revealed on the six individual dimensions of perceived workload. Responses on the postcondition questionnaire also indicated no significant differences between participant's reports of task difficulty and ease of target detection between the different levels of reliability. No participants exhibited notable motion sickness symptoms, and none reported any moderate/severe motion sickness as a result of the test conditions.

5. Discussion

Identifying the advantages of information provided by an automated target detection system is essential in defining design standards for such a system. The objective of this research was to quantify the effects of the percentage of targets missed by an automated system, which provides auditory cues about target location, on target detection performance. The findings of this study will be discussed in light of the hypotheses.

It was hypothesized that as the levels of reliability of the automated system decreased and more targets were missed by the system, the time to detect targets would increase and the percentage of detected targets would decrease. It was also hypothesized that there would be no differences found between the 60% level of reliability and the baseline conditions. The trend of increasing detection time (from 60% to 90%) with increasing reliability was contrary to the hypothesis that detection time would decrease with increasing reliability (i.e., additional sensory cues). The direct relationship between reliability and detection time indicated that the type of cue used in this experiment may not have been conducive to the particular environment. If the targets were less distinguishable from the environment, the results may have shown the inverse relationship

we are accustomed to seeing between reliability and detection time. However, the salience of the target resulted in similar performance between experimental trials with relatively high levels of reliability and trials with no cues at all (i.e., the baseline condition). Another contributor to the direct relationship of reliability and detection time may be the detail provided to each participant regarding the experimental condition that he received. As a result of being provided with this information, participants' performance may have been biased. As the reliability decreased, the priority of the target detection task may have increased. On the contrary, the participants may have prioritized the target detection task less as reliability increased because of their awareness regarding the percentage of cues that they would receive. Also, results that showed the greatest improvement over baseline in the condition with 60% reliability contradicted the findings of Wickens and Dixon (2005), which stated that reliabilities >70% actually provided no improvements over baseline. However, these findings indicated that the cue may have been used as a general audio alert, which increased participant's engagement with the task. Furthermore, the fact that detection time increased with increasing reliabilities indicated that it may have taken longer to interpret the cue than it did to locate the target in the environment alone. As the number of cues the participant interpreted increased, so too did the average detection time per target. With this being the case, the trend of increasing detection time with increasing reliability may be traced to an overreliance on cues which were unsuitable for the task environment.

Also, contrary to the hypothesis and regardless of the reliability level, the relatively high amount of targets identified (98.7 ± 0.4) supported the argument that the targets in this experiment were likely too salient to identify the effects of different reliability levels on target detection reaction time and accuracy. In fact, participants indicated in the postcondition questionnaires that although the target cues were helpful in locating targets, the targets were easy to locate and that despite the reliability level, they looked for targets almost all of the time.

The second hypothesis states that eye movements will increase with each decrease in the level of reliability due to increases in the number of targets missed by the automated system, the frequency of search, and the size of the display area to be searched in absence of cues about target location. Due to extensive amount of effort that went into the development of the simulated task environments, we were unable to configure the eye/head tracking system in a way that would allow us to examine complete range of eye measures needed to fully characterize and analyze eye movement and behavior. However, we were able to conduct an exploratory analysis of the effects of reliability on fixation frequency and duration. These analyses did not support this hypothesis, with no significance differences in fixation frequency or duration between the different levels or reliability. This finding was consistent with the lack of behavioral differences (i.e., accuracy) between the different levels of reliability. Participants did indicate that despite the reliability level, they looked for targets almost all of the time, which may have also contributed to the lack of significant differences in eye-tracking measures among the reliability levels.

Finally, the third hypothesis states that subjective ratings of workload would increase with each decrease in reliability. It was also hypothesized that no significant differences would be found between the 60% level of reliability and the baseline condition in eye movements and subjective ratings of workload. Although the findings contradicted the hypothesis, workload results between the 60%, 75%, and 90% reliability levels reflected corresponding performance results (i.e., no significant difference between reliability levels). This finding was consistent with the common and direct correlation between performance and subjective workload.

6. Conclusions and Future Work

Past research has indicated that automated target acquisition systems augment target detection in a way that facilitates faster detection times and increased target detection (McKinley et al., 1994; Simpson et al., 2004). However, in real-world applications, such target acquisition systems provide imperfect target classification and are prone to type I and type II errors when attempting to identify targets in operational environments. Furthermore, past research indicates a direct relationship between the reliability of the target acquisition system and performance on the target detection task (Wickens et al., 2005). However, few studies address the effect of cue format and target salience on task performance (with respect to reliability level). This experiment highlighted the importance of such context on cueing effectiveness. For this experiment, the salience of the cues paired with nonconducive cue format produced an inverse relationship between reliability and task performance (contradicting findings by previous researchers). Overdependence on these cues, for which interpretation appeared to take longer than target detection alone, resulted in target detection times that actually increased with increasing reliability. Results from this experiment also showed that even infrequent audio tones (as seen in the 60% reliability condition) could potentially increase operator awareness by acting as a general alerting mechanism, resulting in faster target detection times when compared to a baseline condition (i.e., no cue at all). These results provide invaluable information regarding the contextual importance of target detection cues and importance of cue format for target acquisition systems in operational environments.

Additional studies are needed to further explore the advantages of utilizing an automated system to provide target location cues. Future investigations should further investigate the effects of cue format and target salience on target acquisition performance by exploring different types of cue presentations as well as different types and locations of cues within the local environment. And cue presentations shouldn’t be limited to just auditory alerts; future work should also compare the effects of cues presented using different modalities as well as cues presented using different combinations of modalities (e.g., visual and audio). In addition, future experiments should also investigate the benefits of target location cues compared to general attention alerts (i.e., a random

or time-driven tone that occurs throughout a trial), as general alerts are much easier to employ in real-world applications because they aren't dependent on complex target detection systems. Furthermore, since automated target detection systems are also prone to false negative errors, future work should also explore the impact of false alarms (i.e., or providing a cue when a target is not present) on target acquisition performance.

7. References

Beck, H. P.; Dzindolet, M. T.; Pierce, L. G.; Poole, A. H.; McDowell, L. R. Feedback Mitigates Bias Against Machines With Decision-Making Capabilities. Poster session presented at the Annual Meeting of the Southeastern Psychological Association, New Orleans, LA, 2000.

Dzindolet, M. T.; Beck, H. P.; Pierce, L. G. Adaptive Automation: Building Flexibility in Human-Machine Systems. In *Advances in Human Factors and Cognitive Engineering*; Salas, E., Burke, S., Pierce, L., Eds.; Elsevier Press: Boston, 2006; Vol. 6, pp 213–248.

Dzindolet, M. T.; Beck, H. P.; Pierce, L. G.; Dawe, L. A. *A Framework of Automation Use*; ARL-TR-2412; U.S. Army Research Laboratory: Aberdeen Proving Ground, MD, 2001.

Dzindolet, M. T.; Peterson, S. A.; Pomranky, R. A.; Pierce, L. G.; Beck, H. P. The Role of Trust in Automation Reliance. *International Journal of Human Computer Studies* **2003**, *58*, 697–718.

Glumm, M. M.; Kehring, K. L.; White, T. L. *Effects of Tactile, Visual, and Auditory Cues About Threat Location on Target Acquisition and Attention to Visual and Auditory Communications*; ARL-TR-3863; U.S. Army Research Laboratory: Aberdeen Proving Ground, MD, 2006.

Glumm, M. M.; Kehring, K. L.; White, T. L. *Effects of Multimodal Displays About Threat Location on Target Acquisition and Attention to Visual and Auditory Communications*; ARL-TR-4074; U.S. Army Research Laboratory: Aberdeen Proving Ground, MD, 2007.

Hart, S. G.; Staveland, L. E. Development of a Multi-Dimensional Workload Rating Scale: Results of Empirical and Theoretical Research. In *Human Mental Workload*; Hancock, P. A., Meshkati, N., Eds.; Elsevier: Amsterdam, The Netherlands, 1988.

Jacob, R. J. K. Eye-Movement-Based Human-Computer Interaction Techniques: Toward Non-Command Interfaces. In *Advances in Human-Computer Interaction*; Hartson, H. R., Hix, D., Eds.; Ablex Publishing: New Jersey, 1993; Vol. 4, pp 151–190.

Kennedy, R. S.; Lane, N. E.; Lilienthal, M. G.; Berbaum, K. S.; Hettinger, L. J. Profile Analysis of Simulator Sickness Symptoms: Application to Virtual Environment Systems. *Presence* **1992**, *1* (3), 295–370.

McCulloch, C. E. *Generalized Linear Mixed Models*. Institute of Mathematical Statistics: Beachwood, OH, 2003.

McKinley, R. L.; Erickson, M. A.; D'Angelo, W. R. Three-Dimensional Auditory Displays: Development, Applications, and Performance. *Aviation, Space, and Environmental Medicine* **1994**, *5*, A31–A38.

Mitchell, D. K.; Samms, C. L.; Glumm, M. M.; Krausman, A. S.; Brelsford, M. L.; Garrett, L. *Improved Performance Research Integration Tool (IMPRINT) Model Analyses in Support of the Situational Understanding as an Enabler for Unit of Action Maneuver Team Soldiers Science and Technology Objective (STO) in Support of Future Combat Systems (FCS)*; ARL-TR-3405; U.S. Army Research Laboratory: Aberdeen Proving Ground, MD, 2004.

Mosier, K. L.; Skitka, L. J. Human Decision Makers and Automated Decision Aids: Made for Each Other? In *Automation and Human Performance: Theory and Applications*. *Human Factors in Transportation*; Parasuraman, R., Mouloua, M., Eds.; Lawrence Erlbaum Associates: Mahwah, NJ, 1996.

Parasuraman, R.; Molloy, R.; Singh, I. L. Performance Consequences of Automation-Induced "Complacency." *International Journal of Aviation Psychology* **1993**, *3*, 1–23.

Simpson, B. D.; Brungart, D. S.; Gilkey, R. H.; Cowgill, J. L.; Dallman, R. C.; Green, R. F.; Youngblood, K. L.; Moore, T. J. 3-D Audio Cueing for Target Identification in a Simulated Flight Task. *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting*, 2004, pp 1836–1840.

Title 32 Code of Federal Regulations. *Protection of Human Subjects*; Part 219, 1991.

U.S. Department of Defense. *MANPRINT Area Security Analysis Final Report*; Warren, MI, 2005.

U.S. Department of the Army. *Use of Volunteers as Subjects of Research*; Army Regulation (AR) 70-25; Washington, DC, 1990.

Wickens, C. D.; Dixon, S. R. *Is There a Magic Number 7 (to the Minus 1)? The Benefits of Imperfect Diagnostic Automation: A Synthesis of the Literature*; AHFD-05-1/MAAD-05-1;. Aviation Human Factors Division: University of Illinois, Savoy, IL, 2005.

Wickens, C. D.; Dixon, S. R.; Goh, J.; Hammer, B. *Pilot Dependence on Imperfect Diagnostic Automation in Simulated UAV Flights: An Attentional Visual Scanning Analysis*; AHFD-05-2/MAAD-05-2; Aviation Human Factors Division: University of Illinois, Savoy, IL, 2005.

INTENTIONALLY LEFT BLANK.

Appendix A. Demographics Questionnaire

This appendix appears in its original form, without editorial change.

Demographics and Personal Experience Form

1. Participant #: _____ **2. Date: ________________**

3. Age: ______ **4. Handedness: R L** **5. Contacts/Glasses: Y N**

6. Medical Data

a. Are you currently experiencing the effects of any recent illness (cold, flu, etc.) or injury? Yes No
If yes, please list: ____

b. Are you currently taking any cold or flu medications or any anti-motion sickness medications? Yes No
If yes, please list: ____

c. Have you ever experienced moderate to severe Simulator Sickness **(SS)** or Motion Sickness **(MS)**?

SS None/Mild Moderate Severe **MS** None/Mild Moderate Severe

____ Other comments: ____

7. Educational Data

a. What is your highest level of education received?
____GED ____High School ____Some College ____Bachelors Degree ____M.S/M.A
____Ph.D. Other_______________________

b. What subject is your degree in (if applicable, example Engineering) ___________________

8. Military Data (for current military personnel only)

a. Grade:									
	E1	E2	E3	E4	E5	E6	E7	E8	E9
	O1	O2	O3	O4	O5				
	WO1	CWO2	CWO3	CWO4	CWO5				

b. Primary MOS/AFSC: _________ c. Time in MOS/AFSC Years_______ Months______

c. Duty Position/Title: ____

d. Time in present duty position: Years: ______ Months: ______

e. Length of service? Years: ______ Months: ______

9. Prior Experience

a. Do you know how to drive a car or truck? Yes No

b. If so, how many years of experience do you have driving a car or truck? Years Months
A HMMWV? Years _____ Months _____

c. How many hours per week do you use a computer at home or work? Hours _____

d. Do you have any experience with computer games where you control a vehicle (e.g., driving or aircraft simulators)? Yes No
d1. If 'Yes', what games? ____
d2. How many hours per month do you spend playing these games? Hours: _____

e. Have you ever felt dizzy, queasy or disoriented after playing? Yes No

Appendix B. Estimating Simulator Sickness Questionnaire [1]

This appendix appears in its original form, without editorial change.

[1] Kennedy, R. S.; Lane, N. E.; Lilienthal, M. G.; Berbaum, K. S.; Hettinger, L. J. Profile Analysis of Simulator Sickness Symptoms: Application to Virtual Environment Systems. *Presence* **1992**, *1* (3), 295–370.

Participant #:_________ Date:__________________ Time: ____________

Estimating Motion Sickness Questionnaire

Please rate the following measure of motion sickness for the trial performed (or right now) by circling the word that best describes your feelings:

General Discomfort	None	Slight	Moderate	Severe
Fatigue	None	Slight	Moderate	Severe
Headache	None	Slight	Moderate	Severe
Eyestrain	None	Slight	Moderate	Severe
Difficulty Focusing	None	Slight	Moderate	Severe
Increased Salivation	None	Slight	Moderate	Severe
Sweating	None	Slight	Moderate	Severe
Nausea	None	Slight	Moderate	Severe
Difficulty Concentrating	None	Slight	Moderate	Severe
Fullness of Head	None	Slight	Moderate	Severe
Blurred Vision	None	Slight	Moderate	Severe
Dizzy (eyes open)	None	Slight	Moderate	Severe
Dizzy (eyes closed)	None	Slight	Moderate	Severe
Vertigo*	None	Slight	Moderate	Severe
Stomach Awareness**	None	Slight	Moderate	Severe
Burping	None	Slight	Moderate	Severe

* Vertigo is experienced as loss of orientation with respect to vertical upright.

** Stomach awareness is usually used to indicate a feeling of discomfort which is just short of nausea.

Appendix C. Postcondition Questionnaire

This appendix appears in its original form, without editorial change.

Post-Condition Questionnaire

Participant #: ___ Condition ___

Answer each question below by placing an "X" in the bracket that best describes your experience in the experimental condition you *just* completed.

1. How difficult or easy was it to locate targets in the condition that you just completed.

Extremely Difficult	⟵	⟶	*Neither Difficult Nor Easy*	⟵	⟶	*Extremely Easy*
[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]

Comment: ______________________________

2. How would you rate your ability to quickly detect targets in the condition that you just completed?

Poor	⟵	⟶	*Neither Good Nor Bad*	⟵	⟶	*Excellent*
[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]

Comment: ______________________________

If you received target location cues in the last ride, please answer the following questions. Otherwise, stop here.

3. How often did you look for targets that the system may have missed between target location cues?

Never	⟵	⟶	*Half the Time*	⟵	⟶	*All the Time*
[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]

Comment: ______________________________

4. How helpful were the target location cues in finding targets?

Extremely Helpful	⟵	⟶	*Neither Helped Nor Hurt*	⟵	⟶	*Not At All Helpful*
[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]

Comment: ______________________________

Appendix D. National Aeronautics and Space Administration-Task Load Index (NASA-TLX)

This appendix appears in its original form, without editorial change.

RATING SCALE DEFINITIONS

Title	Endpoints	Descriptions
MENTAL DEMAND	Low/High	How much mental and perceptual activity was required (e.g. thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?
PHYSICAL DEMAND	Low/High	How much physical activity was required (e.g. pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
TEMPORAL DEMAND	Low/High	How much time pressure did you feel due to the rate or pace at which the task or task elements occurred? Was the pace slow and leisurely or rapid and frantic?
PERFORMANCE	Perfect/Failure	How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?
EFFORT	Low/High	How hard did you have to work (mentally and physically) to accomplish your level of performance?
FRUSTRATION LEVEL	Low/High	How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

Effort

or

Performance

Temporal Demand

or

Frustration

Temporal Demand

or

Effort

Physical Demand

or

Frustration

Performance

or

Frustration

Physical Demand

or

Temporal Demand

Physical Demand

or

Performance

Temporal Demand

or

Mental Demand

Frustration

or

Effort

Performance

or

Mental Demand

Performance

or

Temporal Demand

Mental Demand

or

Effort

Mental Demand

or

Physical Demand

Effort

or

Physical Demand

Frustration

or

Mental Demand

RATING SCALE

MENTAL DEMAND

Low **High**

PHYSICAL DEMAND

Low **High**

TEMPORAL DEMAND

Low **High**

PERFORMANCE

Good **Poor**

EFFORT

Low **High**

FRUSTRATION

Low **High**

NO. OF COPIES	ORGANIZATION
1 (PDF only)	DEFENSE TECHNICAL INFORMATION CTR DTIC OCA 8725 JOHN J KINGMAN RD STE 0944 FORT BELVOIR VA 22060-6218
1	DIRECTOR US ARMY RESEARCH LAB IMNE ALC HRR 2800 POWDER MILL RD ADELPHI MD 20783-1197
1	DIRECTOR US ARMY RESEARCH LAB RDRL CIM L 2800 POWDER MILL RD ADELPHI MD 20783-1197
1	DIRECTOR US ARMY RESEARCH LAB RDRL CIM P 2800 POWDER MILL RD ADELPHI MD 20783-1197
1	DIRECTOR US ARMY RESEARCH LAB RDRL D 2800 POWDER MILL RD ADELPHI MD 20783-1197
	ABERDEEN PROVING GROUND
1	DIR USARL RDRL CIM G (BLDG 4600)

NO. OF
COPIES ORGANIZATION

1 ARMY RSCH LABORATORY – HRED
RDRL HRM A J MARTIN
MYER CENTER BLDG 2700 RM 2D311
FORT MONMOUTH NJ 07703-5601

1 ARMY RSCH LABORATORY – HRED
RDRL HRM C A DAVISON
320 MANSCEN LOOP STE 115
FORT LEONARD WOOD MO 65473

2 ARMY RSCH LABORATORY – HRED
RDRL HRM DI
T DAVIS
J HANSBERGER
BLDG 5400 RM C242
REDSTONE ARSENAL AL 35898-7290

1 ARMY RSCH LABORATORY – HRED
RDRL HRS EA DR V J RICE
BLDG 4011 RM 217
1750 GREELEY RD
FORT SAM HOUSTON TX 78234-5002

1 ARMY RSCH LABORATORY – HRED
RDRL HRM DG K GUNN
BLDG 333
PICATINNY ARSENAL NJ 07806-5000

1 ARMY RSCH LABORATORY – HRED
ARMC FIELD ELEMENT
RDRL HRM CH C BURNS
THIRD AVE BLDG 1467B RM 336
FORT KNOX KY 40121

1 ARMY RSCH LABORATORY – HRED
AWC FIELD ELEMENT
RDRL HRM DJ D DURBIN
BLDG 4506 (DCD) RM 107
FORT RUCKER AL 36362-5000

1 ARMY RSCH LABORATORY – HRED
RDRL HRM CK J REINHART
10125 KINGMAN RD BLDG 317
FORT BELVOIR VA 22060-5828

1 ARMY RSCH LABORATORY – HRED
RDRL HRM AY M BARNES
2520 HEALY AVE
STE 1172 BLDG 51005
FORT HUACHUCA AZ 85613-7069

NO. OF
COPIES ORGANIZATION

1 ARMY RSCH LABORATORY – HRED
RDRL HR MP D UNGVARSKY
POPE HALL BLDG 470
BCBL 806 HARRISON DR
FORT LEAVENWORTH KS 66027-2302

1 ARMY RSCH LABORATORY – HRED
RDRL HRM DQ M R FLETCHER
NATICK SOLDIER CTR
AMSRD NSC WS E BLDG 3 RM 343
NATICK MA 01760-5020

1 ARMY RSCH LABORATORY – HRED
RDRL HRM AT J CHEN
12350 RESEARCH PKWY
ORLANDO FL 32826-3276

1 ARMY RSCH LABORATORY – HRED
RDRL HRM AT C KORTENHAUS
12350 RESEARCH PKWY
ORLANDO FL 32826

1 ARMY RSCH LABORATORY – HRED
RDRL HRM AS C MANASCO
SIGNAL TOWERS
BLDG 29808A RM 303A
FORT GORDON GA 30905-5233

1 ARMY RSCH LABORATORY – HRED
RDRL HRM CU
6501 E 11 MILE RD MS 284
BLDG 200A 2ND FL RM 2104
WARREN MI 48397-5000

1 ARMY RSCH LABORATORY – HRED
FIRES CTR OF EXCELLENCE
FIELD ELEMENT
RDRL HRM AF C HERNANDEZ
3040 NW AUSTIN RD RM 221
FORT SILL OK 73503-9043

1 ARMY RSCH LABORATORY – HRED
RDRL HRM AV S MIDDLEBROOKS
91012 STATION AVE RM 348
FORT HOOD TX 76544-5073

1 ARMY RSCH LABORATORY – HRED
RDRL HRM CN R SPENCER
DCSFDI HF
HQ USASOC BLDG E2929
FORT BRAGG NC 28310-5000

NO. OF COPIES	ORGANIZATION
1	ARMY RSCH LABORATORY – HRED RDRL HRM DW E REDDEN BLDG 4 CL 60 FORT BENNING GA 31905-5400
1 (CD only)	ARMY G1 DAPE MR B KNAPP 300 ARMY PENTAGON RM 2C489 WASHINGTON DC 20310-0300
	ABERDEEN PROVING GROUND
5	DIR USARL RDRL CIM G S FOPPIANO RDRL HR L ALLENDER T LETOWSKI RDRL HRM B J LOCKETT RDRL HRS D B AMREIN

NO. OF COPIES	ORGANIZATION
	ABERDEEN PROVING GROUND
10	DIR USARL RDRL HRS B T WHITE (5 CPS) J DAVIS (5 CPS)

INTENTIONALLY LEFT BLANK.